DOCUMENT RESUME

ED 435 692 TM 030 323

AUTHOR van Engelenburg, Gijsbert
TITLE Statistical Analysis for the Solomon Four-Group Design. Research Report 99-06.
INSTITUTION Twente Univ., Enschede (Netherlands). Faculty of Educational Science and Technology.
PUB DATE 1999-00-00
NOTE 38p.
AVAILABLE FROM Faculty of Educational Science and Technology, University of Twente, TO/OMD, P.O. Box 217, 7500 AE Enschede, The Netherlands.
PUB TYPE Numerical/Quantitative Data (110) -- Reports - Descriptive (141)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Foreign Countries; *Maximum Likelihood Statistics; *Outcomes of Treatment; *Pretests Posttests; *Regression (Statistics)
IDENTIFIERS *Solomon Four Group Design

ABSTRACT

The Solomon four-group design (R. Solomon, 1949) is a very useful experimental design to investigate the main effect of a pretest and the interaction of pretest and treatment. Although the design was proposed half a century ago, no proper data analysis techniques have been available. This paper describes how data from the Solomon four-group design can be analyzed properly using maximum likelihood regression analysis. The results show that both the parameter estimators and the expressions for the standard errors are in agreement with related designs and with intuitive expectations. (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 435 692

TM

Statistical Analysis for the Solomon Four-Group Design

Research Report 99-06

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

J. Nelissen

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Bert van Engelenburg

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM030323

faculty of
EDUCATIONAL SCIENCE AND TECHNOLOGY

University of Twente

Department of Educational Measurement and Data Analysis

BEST COPY AVAILABLE

Statistical Analysis
for the Solomon Four-Group Design

Gijsbert van Engelenburg

Abstract

The Solomon four-group design is a very useful experimental design to investigate the main effect of a pretest and the interaction effect of pretest and treatment. Although the design has been proposed half a century ago, no proper data analysis techniques have been available. In this paper it is described how data from the Solomon four-group design can be properly analyzed using maximum likelihood regression analysis. The results show that both the parameter estimators and the expressions for the standard errors are in agreement with related designs and with intuitive expectations.

Statistical analysis for the Solomon Four-Group Design

Whether pretests should be administered in experimental control group designs is a delicate issue. On the one hand, the statistical argument is that pretests are usually favorable, since controlling for pretest scores will usually increase the power of the design by reducing residual variance. If a pretest is administered, the probability that small treatment effects are detected will usually increase. On the other hand, the methodological argument is that pretests are often unfavorable, since pretests can blur the real treatment effect. Two situations must be distinguished (see e.g. Campbell & Stanley, 1966; Cook & Campbell, 1979). First, the pretest can have an effect by itself by decreasing or increasing the scores of interest. For instance, administering a pretest may offer the subject additional exercise material, item training or a search strategy. Such an effect, commonly referred to as a *testing* effect, is a main effect of the pretest. Second, the pretest can make the treatment less or more effective. For instance, administering a pretest may arouse curiosity or it may sensitize the subject. Such an effect is an interaction effect of pretest and treatment. Although we will usually not expect that pretest main effects and interaction effects are very large (see Wilson & Putnam, 1982), it will be usually difficult to predict whether the statistical or the methodological argument prevails in a given situation.

To investigate the role of a pretest the Solomon four-group design (Solomon, 1949) is a useful device. Using this design the treatment effect, the pretest effect, and the interaction effect of pretest and treatment can be disentangled. A graphical depiction of the design is shown in Figure 1. Subjects are randomly (denoted by R) assigned to one of four groups (denoted by O, P, T, and TP), presented rowwise. Columnwise the experimental interventions are presented. Group P and TP get a pretest (X), group T and TP get the treatment (T), and all groups get a posttest (Y).

Insert Figure 1 about here

A striking impression from the existing literature is that the same authors who praise the Solomon four-group design as a very useful device, admit that the statistical treatment of the design is not straightforward. Oliver and Berger (1980) studied several pretest designs and concluded that only the Solomon four-group design

yielded unambiguous results. They also noted that there are many possible ways of analyzing the data, which is in this case another way of saying that no single proper analysis technique is known. Solomon's (1949) suggestions are not of much help and even "judged unacceptable" by Campbell and Stanley (1966, p. 25). They noted that "[t]here is no singular statistical procedure which makes use of all six sets of observations simultaneously" (p.25). Sawilovski, Kelly, Blair, and Markman (1994) believe that the lack of a clear statistical treatment is probably the reason that there are few examples of published research based on the Solomon four-group design.

The most straightforward statistical treatment is probably to reorganize the posttest data according to a 2×2 factorial design, where the first factor is whether a pretest has been given or not, and the second whether a treatment has been given or not (Figure 2). It is clear that an ANOVA can provide estimates of the treatment effect, of the pretest effect, and of the interaction effect of pretest and treatment. However, by analyzing the data this way, not all available information is used. Administering pretests will, of course, result in pretest scores. Although pretest scores could be used to obtain better estimates, they are neglected in the ANOVA model. Campbell & Stanley (1966) advise to use the 2×2 ANOVA design and they conclude that "[i]f the main and interactive effects of pretesting are negligible, it may be desirable to perform an analysis of covariance..." (Campbell & Stanley, 1966, p.25), where they refer to analyzing only the data from group P and TP to test whether an effect of the treatment can be found. (See also Huck & Sandler, 1973.)

Insert Figure 2 about here

A second obvious way to analyse the data from the Solomon four-group design is more or less in the spirit of the latter advice of Campbell and Stanley (1966). It is possible to view the Solomon four-group as a combination of two experiments. In the first experiment only group O and T are considered. Since no pretest was administered, this experiment can be treated as a posttest-only control group experimental design. In the second experiment only group P and TP are considered. In these groups a pretest was administered and this experiment can be treated as a pretest-posttest control group experimental design.

Some other attempts are made to analyse Solomon four-group data. Walton Braver & Braver (1988) have suggested the use of a meta-analytic approach to combine the results of both separate experiments. Their approach has been criticized by Sawilovski, Kelly, Blair, & Markman (1994), who show that the approach results in a lack of power and an inflation of the nominal alpha. They conclude that "further study is needed to resolve the data analysis problem associated with this design" (Sawilovski, Kelly, Blair, & Markman, 1994, p 374). Dukes, Ullman, and Stein (1995) made multiple comparisons between two groups at the time using a latent variable approach.

In this paper a full information maximum likelihood method is presented to analyze all available data from the Solomon four-group design. The results will be compared with the results of the 2×2 factorial design and with the results of the combination of the pretest-only and the pretest-posttest designs. First, these commonly used approaches will be presented, then the new approach. All approaches will be presented in a maximum likelihood regression context, to allow simple comparisons between them. Before presenting the models, however, some general remarks are made, concerning notation, assumptions and estimation procedure.

General remarks

Notation

First some notation will be introduced. The four groups of the Solomon four-group design will be denoted by Q = O, P, T, TP, as indicated in Figure 2. Groups will be treated as nonoverlapping sets of subjects. The four groups are identified by two dichotomous design variables. The first design variable, T_i, indicates whether subject i has received a treatment ($T_i = 1$) or not ($T_i = 0$) and, second, P_i is the design variable which indicates whether for subject i a pretest has been administered ($P_i = 1$) or not ($P_i = 0$).

Subjects are denoted by $i = 1, 2, ..., N$, where N is the total number of subjects. The number of subjects in group Q will be indicated by N_Q. The posttest score of subject i will be denoted by Y_i, the pretest score, if available, by X_i. In this paper summation is always over subjects and the summation index denotes the group in

which summation takes place. If more than one summation index is given, summation takes place in all groups mentioned. Thus

$$\sum_{Q,R}(Y_i) = \sum_{Q}(Y_i) + \sum_{R}(Y_i) = \sum_{i\in Q}(Y_i) + \sum_{i\in R}(Y_i). \tag{1}$$

is the sum of the posttest scores of all subjects belonging to group Q and R.

Further, many formulas contain sum of squares or sums of crossproducts and it is useful to have a shorthand notation for them. We will define the function SST_Q to be

$$SST_Q(U,V) = \sum_{Q} U_i V_i \tag{2}$$

which is the "total" sum of crossproducts within a single group or, if $U = V$, the sum of squares. Further, we will define the function SSW_Q to be

$$SSW_Q(U,V) = \sum_{i\in Q}(U_i - \frac{1}{N_Q}\sum_{j\in Q} U_j)(V_i - \frac{1}{N_Q}\sum_{j\in Q} V_j), \tag{3}$$

which is the "within" group sum of crossproducts or sum of squares. Finally, the function SSB_Q is defined as

$$SSB_Q(U,V) = \frac{1}{N_Q}\sum_{Q} U_i \sum_{Q} V_i, \tag{4}$$

which is the "between" group sum of crossproducts or sum of squares. It is well known and easy to verify that

$$SST_Q(U,V) = SSW_Q(U,V) + SSB_Q(U,V). \tag{5}$$

Multiple indices will be used to denote summation over several groups. For example,

$$SSW_{Q.R}(U,V) = SSW_{Q}(U,V) + SSW_{R}(U,V), \tag{6}$$

which is generally unequal to $SSW_{Q\cup R}(U,V)$. An analogous result is true for $SSB_{Q.R}(U,V)$.

Assumptions

To allow maximum likelihood estimation, it will be assumed that subjects are randomly assigned to the four groups. For the sake of generality, we will not assume that all groupsizes are equal. Further, we will assume that error components are independently and normally distributed with mean zero and fixed, but unknown variance. It will also be assumed that the size of all experimental effects are fixed, but unknown. Thus, first, the size of a potential treatment effect is equal for all subjects who have received the experimental treatment, second, the size of a potential (main) effect of the pretest is also equal for all subjects who have received the pretest, and, third, the size of a potential interaction effect is also equal for all subjects who received both a pretest and a treatment.

Further, as far as pretested subjects are considered, a linear relation is assumed between pretest scores and posttest scores. For notational convenience we will assume that pretest scores, but not posttest scores, are deviation scores from the observed mean. Thus, if X'_i is the raw pretest score and

$$\overline{X} = \frac{1}{N_P + N_{TP}} \sum_{i \in P \cup TP} X'_i, \tag{7}$$

then

$$X_i = X'_i - \overline{X}. \tag{8}$$

It follows that

$$\sum_{P} X_i = -\sum_{TP} X_i, \tag{9}$$

which will be useful in simplifying some of the formulas.

Estimation

The models presented in the remainder were all analyzed in the same way, according to maximum likelihood estimation. Maximum likelihood estimation for regression models is the relatively straightforward. Here, estimation of the parameters and their standard errors is described only briefly. First, the likelihood of the data is expressed as a function of the model parameters, given the observed data. Then, first derivatives of the logarithm of the likelihood function to the parameters are algebraically derived. The first derivatives are set equal to zero and from this set of equations the parameters must be solved. For the models below, this proves to be not too difficult. To obtain estimates of the standard errors, first the matrix of second derivatives of the logarithm of the likelihood function to the parameters must be obtained. Next, the negative of this matrix must be inverted. The matrix, however, proves to be relatively simple and inverting can be done algebraically by standard mathematics software. The inverted matrix is the variance-covariance matrix of the maximum likelihood parameter estimates. Standard errors can be computed by taking the square root of the appropriate diagonal elements.

Commonly used methods

The 2 × 2 factorial model

If the Solomon four-group design is analysed as a 2 × 2 factorial design, the pretest scores are ignored. We are, of course, interested in differences between the four groups in the mean value of Y_i and the model is

$$Y_i = a + b_T T_i + b_P P_i + b_{TP} T_i P_i + R_i, \qquad (10)$$

where b_T, b_P, and b_{TP} are the regression coefficients associated with T_i, P_i, and the interaction effect of T_i and P_i, respectively and R_i is the residual term. The coefficient

b_T is interpreted as the treatment effect, b_P is interpreted as the (main) effect of the pretest, and b_{TP} is interpreted as the additional effect if both a treatment and a pretest are given.

Many textbooks on regression analysis or analysis of variance treat the estimation of parameters of a 2 × 2 factorial model. Here, maximum likelihood estimation is applied to allow comparisons between this model and the other models. If it is assumed that R_i is independently normally distributed with mean zero and variance σ_R^2,

$$R_i \sim N(0,\sigma_R^2), \tag{11}$$

then it follows that Y_i is independently normally distributed with mean $a + b_T T_i + b_P P_i + b_{TP} T_i P_i$ and variance σ_R^2, so that the logarithm of the likelihood function is

$$logL = -\frac{N}{2}\log(2\pi) - N\log\sigma_R - \frac{1}{2}\sum\left(\frac{Y_i - a - b_T T_i - b_P P_i - b_{TP} T_i P_i}{\sigma_E}\right)^2, \tag{12}$$

where summation is over all subjects. Parameter estimation is without problems and the resulting parameter estimates are given in Table 1. Also, the variance of the parameter estimates are not too difficult to find. They are also presented in Table 1.

Insert Table 1 about here

Using the 2 × 2 factorial model to analyze Solomon four-group data pretest scores are neglected. We may expect that parameter estimates can be improved by using pretest scores. Also, a reduction of the standard errors of the estimates can be expected.

The posttest-only and pretest-posttest model

The Solomon four-group design can be conceived as two separate experiments, a posttest-only control group experiment and an experiment with pretest and posttest. Below it is investigated what the result will be if both models are analyzed separately.

For the subjects in groups O and T, who did not have a pretest, the model simply is

$$Y_i = a + b_T T_i + R_i, \qquad \text{for } i \in O \cup T. \tag{13}$$

For this model parameter estimation is very much analogous to the estimation for the 2×2 factorial design. It follows that Y_i is independently normally distributed with mean $a + b_T T_i$ and variance σ_R^2, so that the logarithm of the likelihood function is

$$logL = -\frac{N_O + N_T}{2}\log(2\pi) - (N_O + N_T)\log\sigma_R - \frac{1}{2}\sum_{O,T}\left(\frac{Y_i - a - b_T T_i}{\sigma_R}\right)^2. \tag{14}$$

The maximum likelihood parameter estimators are presented in Table 2, as are the associated variance estimates.

Insert Table 2 about here

The subjects in groups P and TP, however, did have a pretest and pretest scores can be entered into the regression analysis to reduce the residual variance. Since all subjects did have a pretest the *main* effect of the pretest can not be separated from the intercept. This combined parameter will be denoted by a', thus

$$a' = a + b_P. \tag{15}$$

Also, in this model an effect of the treatment cannot be separated from the interaction effect of pretest and treatment. This combined effect will be denoted by b_T', thus

$$b_T' = b_T + b_{TP}. \tag{16}$$

The model, then, is

$$Y_i = a' + b_T' T_i + b_X X_i + E_i, \qquad \text{for } i \in P \cup PT, \qquad (17)$$

where, b_X is the regression coefficient associated with X_i. The error term E_i will usually be smaller than the residual terms R_i in (13), because X_i is now entered into the model.

To find the parameters of this model, it is assumed that E_i is independently normally distributed with mean zero and variance σ_E^2,

$$E_i \sim N(0, \sigma_E^2). \qquad (18)$$

It follows that Y_i is independently normally distributed with mean $a' + b_T' T_i + b_X X_i$ and variance σ_E^2 and the logarithm of the likelihood function is

$$logL = -\frac{N_P + N_{TP}}{2}\log(2\pi) - (N_P + N_{TP})\log\sigma_E - \frac{1}{2}\sum_{P,TP}\left(\frac{Y_i - a' - b_T' T_i - b_X X_i}{\sigma_E}\right)^2 \qquad (19)$$

and, again, parameters and standard errors can be obtained in the same way. In Table 3 parameter estimates and associated variance estimates are presented.

Insert Table 3 about here

In this approach not all parameters of interest are estimated explicitly. However, the remaining parameters can be obtained by combining the results of Table 2 and 3. Because of (15) an estimate of b_P can be obtained by

$$\hat{b}_P = \hat{a}' - \hat{a} = \frac{\sum_P Y_i - \hat{b}_X \sum_P X_i}{N_P} - \frac{\sum_O Y_i}{N_O} \qquad (20)$$

and in the same way, because of (16) an estimate of b_{TP} can be obtained by

$$\hat{b}_{TP} = \hat{b}'_T - \hat{b}_T = \frac{\sum_{TP} Y_i - \hat{b}_X \sum_{TP} X_i}{N_{TP}} - \frac{\sum_{P} Y_i - \hat{b}_X \sum_{P} X_i}{N_P} - \frac{\sum_{T} Y_i}{N_T} + \frac{\sum_{O} Y_i}{N_O} \qquad (21)$$

These results have not been stated explicitly before, although many researchers probably had the intuitive notion that combining the results of both designs gave insight about the pretest (main) effect and the interaction effect of pretest and treatment. An important question, however, remains. Although the estimators are maximum likelihood estimators, it may be doubted whether they are efficient. Limited information estimators, which result if not all available data are analyzed simultaneously, may be less efficient than full information estimators. The variance of $\hat{b}_P$ and the variance of $\hat{b}_{TP}$ will not be derived here. After the new approach is presented, a remark regarding their respective variances will be made.

There are two main reasons to use pretest scores, and both can be illustrated with the current parameter estimates at hand. First, using pretest scores may correct for sampling fluctuations. What we see in Table 3 is that some of the parameter estimates are not simply the mean of the posttest scores within a certain group, but that the mean score is corrected for pretest differences. Second, using pretest scores will usually result in a smaller residual term and thereby enhancing the power of the significance tests of the parameters. Indeed, from $\hat{\sigma}_E$ it is clear that the sum of squares of posttest scores corrected for pretest differences will be usually much smaller than simply the sums of squares.

New method

Model

The main flaws of the previous approach results from analyzing both models separately. Below is presented how the data from all four groups, including pretest data, can be analyzed simultaneously. If pretest scores were available for all subjects, the model is defined as

$$Y_i = a + b_X X_i + b_T T_i + b_P P_i + b_{TP} T_i P_i + E_i, \qquad (22)$$

To avoid the problem of the structurally missing pretest scores, for each of the four groups a separate model will be derived from (22). First, for the subjects who did have both a pretest and a treatment the model can be written as

$$Y_i = a + b_X X_i + b_T + b_P + b_{TP} + E_i, \qquad \text{for } i \in \text{TP}. \tag{23}$$

Second, for the subjects who did have a pretest, but no treatment the model is

$$Y_i = a + b_X X_i + b_P + E_i, \qquad \text{for } i \in \text{P}. \tag{24}$$

If no pretest is administered, and thus, no pretest scores are available, the variation that X_i accounts for is added to the residual variation. If the residual part is denoted by R_i, the model for the subjects that did have a treatment but no pretest can be written as

$$Y_i = a + b_T + R_i, \qquad \text{for } i \in \text{T}. \tag{25}$$

and for the subjects that had neither a pretest nor a treatment the model is

$$Y_i = a + R_i, \qquad \text{for } i \in \text{O}. \tag{26}$$

Conceptually, the residual part R_i can be decomposed into two parts, the "real" error term and a term that is related to pre-experimental differences,

$$R_i = b_X X_i + E_i, \tag{27}$$

although X_i is unobserved for these subjects. Note that, although a separate model is formulated for each group, they will be analyzed simultaneously.

Parameter estimation

Since the parameter estimation for this new model has not been described before, a more detailed desciption will be given. From (23) to (26) and the assumption of independently, normally distributed error terms, the distribution of Y_i can be found. For each group it follows that Y_i is normally distributed, with means and standard deviation depending on the group,

$$f(Y_i) = \frac{1}{\sigma_R\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{Y_i - a}{\sigma_R}\right)^2\right\}, \quad \text{for } i \in \text{O}, \quad (28)$$

$$f(Y_i) = \frac{1}{\sigma_R\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{Y_i - a - b_T}{\sigma_R}\right)^2\right\}, \quad \text{for } i \in \text{T}, \quad (29)$$

$$f(Y_i) = \frac{1}{\sigma_E\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{Y_i - a - b_X X_i - b_P}{\sigma_E}\right)^2\right\}, \quad \text{for } i \in \text{P}, \quad (30)$$

and

$$f(Y_i) = \frac{1}{\sigma_E\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{Y_i - a - b_X X_i - b_T - b_P - b_{TP}}{\sigma_E}\right)^2\right\}, \quad \text{for } i \in \text{TP}. \quad (31)$$

Since the likelihood of the parameters given the observed data is

$$L(a, b_X, b_T, b_P, b_{TP}, \sigma_R, \sigma_E \mid Y_i) = \prod_{i=1}^{N} f(Y_i), \quad (32)$$

the logarithm of the likelihood, written in terms separately for each group, is

$$logL = \sum_{O}\left(\log\frac{1}{\sqrt{2\pi}} + \log\frac{1}{\sigma_R} - \frac{1}{2}\left(\frac{Y_i - a}{\sigma_R}\right)^2\right)$$

$$+ \sum_{T}\left(\log\frac{1}{\sqrt{2\pi}} + \log\frac{1}{\sigma_R} - \frac{1}{2}\left(\frac{Y_i - a - b_T}{\sigma_R}\right)^2\right)$$

$$+ \sum_{P}\left(\log\frac{1}{\sqrt{2\pi}} + \log\frac{1}{\sigma_E} - \frac{1}{2}\left(\frac{Y_i - a - b_X X_i - b_P}{\sigma_E}\right)^2\right)$$

$$+ \sum_{TP}\left(\log\frac{1}{\sqrt{2\pi}} + \log\frac{1}{\sigma_E} - \frac{1}{2}\left(\frac{Y_i - a - b_X X_i - b_T - b_P - b_{TP}}{\sigma_E}\right)^2\right). \qquad (33)$$

This can be simplified to

$$logL = -N\log\sqrt{2\pi} - (N_O + N_T)\log\sigma_R - (N_P + N_{TP})\log\sigma_E$$

$$- \frac{1}{2\sigma_R^2}\sum_{O,T} Y_i^2 - \frac{1}{2\sigma_E^2}\sum_{P,TP} Y_i^2 + \frac{a}{\sigma_R^2}\sum_{O} Y_i + \frac{a + b_T}{\sigma_R^2}\sum_{T} Y_i$$

$$+ \frac{a + b_P}{\sigma_E^2}\sum_{P} Y_i + \frac{a + b_P + b_T + b_{TP}}{\sigma_E^2}\sum_{TP} Y_i - \frac{b_X^2}{2\sigma_E^2}\sum_{P,TP} X_i^2 - \frac{b_X(a + b_P)}{\sigma_E^2}\sum_{P} X_i$$

$$- \frac{b_X(a + b_P + b_T + b_{TP})}{\sigma_E^2}\sum_{TP} X_i + \frac{b_X}{\sigma_E^2}\sum_{P,TP} Y_i X_i - N_O\frac{a^2}{2\sigma_R^2} - N_T\frac{(a + b_T)^2}{2\sigma_R^2}$$

$$- N_P\frac{(a + b_P)^2}{2\sigma_E^2} - N_{TP}\frac{(a + b_T + b_P + b_{TP})^2}{2\sigma_E^2}. \qquad (34)$$

The first derivatives are presented in Appendix 1. Setting the equations equal to zero results in the parameter estimators in Table 4. This exercise, although not difficult is somewhat space and time consuming.

Insert Table 4 about here

Standard errors

The second derivatives are presented in Appendix 2. The negative of the matrix of second derivatives must be inverted to obtain the variance-covariance matrix.

Fortunately, the expressions for the matrix of second derivatives are relatively simple and since both the second derivatives to σ_E and any other parameter and the second derivatives to σ_R and any other parameter are all zero, only a 5×5 matrix has to be inverted. Algebraic invertion is not problematic using standard mathematics software. The expressions for the variances of the parameter estimates are presented in Table 4, for the covariances in Table 5.

Insert Table 5 about here

Discussion

The investigation for a new approach was motivated by the thought that by using pretest scores better results would be likely than that would be obtained by using the 2×2 factorial design. This means, at least, two different things. First, since sampling fluctuations can play a role, it is advisable to correct for pre-experimental differences if possible. Second, by reducing the residual variance, standard errors of the estimates will be smaller and, thus, significance tests more powerfull. Below we will discuss whether the new model fulfils the expectations.

First a comparison with the 2×2 factorial model will be made. If the estimators of a and b_T for the new model (see Table 4) are compared with the estimators for the factorial model (see Table 1) it is clear that both the estimators and the standard errors are exactly the same for both models. Undoubtly, this has to do with the fact that these estimators are only based on the data of the groups that were not pretested. But, for b_P and b_{TP} there are differences. The estimators show that the posttest scores in groups P and TP are indeed corrected for pretest differences. Variance estimates are changed too. Some σ_R terms are replaced by σ_E terms and, since σ_E will usually be smaller, often much smaller, than σ_R, this indeed indicates that the variance estimates are shrunken. However, this is not the whole story. A component consisting of the ratio of "between" sums of squares of X_i and "within" sums of squares of X_i is added to the σ_E terms as a kind of penalty. Usually, this ratio will be very small, but it reminds us that adding a pretest may even work out worse in extreme cases. If X_i is totally unrelated to Y_i, σ_E will be equal to σ_R and the result is that the variance estimates of b_P and b_{TP} will be even larger for the new model than for the factorial model.

Comparison to the combination of the posttest-only and the pretest-posttest model yield a different picture. If the estimates of a and b_T for the new model (see Table 4) are compared with the estimates for the combination of the pretest-only model (see Table 2) and the pretest-posttest model (see Table 3) it is clear that, again for a and b_T, there is full agreement between the models. As has been shown in (20) and (21) b_P and b_{TP} can be estimated by combining the results of the pretest-only model and the pretest-posttest model. It is clear that they are exactly equal to the parameter estimates of the new model. However, for the new model standard errors were directly obtained, which was not the case for the combination of the pretest-only

model and the pretest-posttest model. It is interesting to see that $Var(\hat{b}_P) = Var(\hat{a}') + Var(\hat{a})$ and that $Var(\hat{b}_{TP}) = Var(\hat{b}'_T) + Var(\hat{b}_T)$. In retrospect, this is a plausible result. This result could have been derived from (20) and (21) by realizing that $Cov(\hat{a}', \hat{a}) = 0$ and $Cov(\hat{b}'_T, \hat{b}_T) = 0$.

In this study it has been assumed that both E_i and R_i are normally distributed. This implies that X_i is also normally distributed (see (27)). These assumpties allow maximum likelihood estimation. However, the assumptions may not allways be realistic and, therefore, in some cases least squares estimation should be preferred. There is another reason to prefer least squares estimation. It is well known that the maximum likelihood estimates of the residual variance are somewhat biased, although the bias reduces with larger sample size. Maximum likelihood estimates of the residual variance can be corrected to remove the bias by adjusting the number of degrees of freedom and in that case they are equal to least squares estimates. Although not investigated, a suggestion is to remove the bias of $\hat{\sigma}_R$ by dividing by $N_O + N_T - 2$ instead of by $N_0 + N_T$ and to remove the bias of $\hat{\sigma}_E$ by dividing by $N_P + N_{TP} - 3$ instead of $N_P + N_{TP}$ (see Table 4). This will probably yield the least squares estimates, but that topic needs further investigation.

A more or less surprising result of this study is that the results of the new model could have been found by combining the results of the pretest-only model and the pretest-posttest model. Since those models are very well known, it is a bit striking that so many authors have mentioned that no proper analysis technique existed for the Solomon four-group design. However, the findings in this paper will hopefully result in more studies that implement this very useful design.

References

Campbell, D.T. & Stanley, J.C. (1966). *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally.

Cook, T.D. & Campbell, D.T. (1979). *Quasi-Experimentation: Design and analysis issues for field settings.* Boston: Houghton Miffin.

Dukes, R.L., Ullman, J.B., & Stein, J.A. (1995). An evaluation of DARE (Drug-abuse resistance education), using a Solomon 4-group design with latent-variables. *Evaluation Review, 19* (4), 409-435.

Huck, S.W., & Sandler, H.M. (1973). A note on the Solomon 4-group design: Appropriate statistical analyses. *The Journal of Experimental Education, 42*, 54-55.

Oliver, R.L. & Berger, P.K. (1980). Advisability of pretest designs in psychological research. *Perceptual and Motor Skills, 51*, 463-471.

Sawilowski, S., Kelly, D.L., Blair, R.C., & Markman, B.S. (1994). Meta-analysis and the Solomon four-group design. *Journal of Experimental Education, 62* (4), 361-376.

Solomon, R.L. (1949). An extension of control group design. *Psychological Bulletin, 46*, 137-150.

Walton Braver, M.C. & Braver, S.L. (1988). Statistical treatment of the Solomon four-group design: A meta-analytic approach. *Pychological Bulletin, 104*, 150-154.

Wilson, V.L. & Putnam, R.R. (1982). A meta-analysis of pretest sensitization effects in experimental design. *American Educational Research Journal, 19*, 249-258.

Table 1. Parameter estimators for the 2×2 factorial model.

θ	$\hat{\theta}$	$Var(\hat{\theta})$
a	$\frac{\sum_{O} Y_i}{N_O}$	$\frac{\hat{\sigma}_R^2}{N_O}$
b_T	$\frac{\sum_{T} Y_i}{N_T} - \frac{\sum_{O} Y_i}{N_O}$	$\frac{\hat{\sigma}_R^2}{N_O} + \frac{\hat{\sigma}_R^2}{N_T}$
b_P	$\frac{\sum_{P} Y_i}{N_P} - \frac{\sum_{O} Y_i}{N_O}$	$\frac{\hat{\sigma}_R^2}{N_O} + \frac{\hat{\sigma}_R^2}{N_P}$
b_{TP}	$\frac{\sum_{TP} Y_i}{N_{TP}} - \frac{\sum_{T} Y_i}{N_T} - \frac{\sum_{P} Y_i}{N_P} + \frac{\sum_{O} Y_i}{N_O}$	$\frac{\hat{\sigma}_R^2}{N_O} + \frac{\hat{\sigma}_R^2}{N_T} + \frac{\hat{\sigma}_R^2}{N_P} + \frac{\hat{\sigma}_R^2}{N_{TP}}$
σ_R	$\sqrt{\frac{SSW_{O,T,P,TP}(Y,Y)}{N}}$	$\frac{\hat{\sigma}_R^2}{2N}$

Table 2. Parameter estimators for the posttest-only model.

θ	$\hat{\theta}$	$Var(\hat{\theta})$
a	$\frac{\sum_{O} Y_i}{N_O}$	$\frac{\hat{\sigma}_R^2}{N_O}$
b_T	$\frac{\sum_{T} Y_i}{N_T} - \frac{\sum_{O} Y_i}{N_O}$	$\frac{\hat{\sigma}_R^2}{N_O} + \frac{\hat{\sigma}_R^2}{N_T}$
σ_R	$\sqrt{\frac{SSW_{O,T}(Y,Y)}{N_O + N_T}}$	$\frac{\hat{\sigma}_R^2}{2(N_O + N_T)}$

Table 3. Parameter estimators for the pretest-posttest model.

θ	$\hat{\theta}$	$Var(\hat{\theta})$
b_X	$\frac{SSW_{P,TP}(Y,X)}{SSW_{P,TP}(X,X)}$	$\frac{\sigma_E^2}{SSW_{P,TP}(X,X)}$
a' $(=a+b_P)$	$\frac{\sum_P Y_i - \hat{b}_X \sum_P X_i}{N_P}$	$\frac{\hat{\sigma}_E^2}{N_P}\left(1+\frac{SSB_P(X,X)}{SSW_{P,TP}(X,X)}\right)$
b_T' $(=b_T+b_{TP})$	$\frac{\sum_{TP} Y_i - \hat{b}_X \sum_{TP} X_i}{N_{TP}} - \frac{\sum_P Y_i - \hat{b}_X \sum_P X_i}{N_P}$	$\left(\frac{\hat{\sigma}_E^2}{N_P}+\frac{\hat{\sigma}_E^2}{N_{TP}}\right)\left(1+\frac{SSB_{P,TP}(X,X)}{SSW_{P,TP}(X,X)}\right)$
σ_E	$\sqrt{\frac{SSW_{P,TP}(Y_i-\hat{b}_X X_i, Y_i-\hat{b}_X X_i)}{N_P+N_{TP}}}$	$\frac{\hat{\sigma}_E^2}{2(N_P+N_{TP})}$

Table 4. Parameter estimators for the new model.

θ	$\hat{\theta}$	$Var(\hat{\theta})$
b_X	$\frac{SSW_{P,TP}(Y,X)}{SSW_{P,TP}(X,X)}$	$\frac{\sigma_E^2}{SSW_{P,TP}(X,X)}$
a	$\frac{\sum_O Y_i}{N_O}$	$\frac{\hat{\sigma}_R^2}{N_O}$
b_T	$\frac{\sum_T Y_i}{N_T} - \frac{\sum_O Y_i}{N_O}$	$\frac{\hat{\sigma}_R^2}{N_O} + \frac{\hat{\sigma}_R^2}{N_T}$
b_P	$\frac{\sum_P Y_i - \hat{b}_X \sum_P X_i}{N_P} - \frac{\sum_O Y_i}{N_O}$	$\frac{\hat{\sigma}_E^2}{N_P}\left(1+\frac{SSB_P(X,X)}{SSW_{P,TP}(X,X)}\right) + \frac{\hat{\sigma}_R^2}{N_O}$
b_{TP}	$\frac{\sum_{TP} Y_i - \hat{b}_X \sum_{TP} X_i}{N_{TP}} - \frac{\sum_P Y_i - \hat{b}_X \sum_P X_i}{N_P} - \frac{\sum_T Y_i}{N_T} + \frac{\sum_O Y_i}{N_O}$	$\left(\frac{\hat{\sigma}_E^2}{N_P} + \frac{\hat{\sigma}_E^2}{N_{TP}}\right)\left(1+\frac{SSB_{P,TP}(X,X)}{SSW_{P,TP}(X,X)}\right) + \frac{\hat{\sigma}_R^2}{N_O} + \frac{\hat{\sigma}_R^2}{N_T}$
σ_R	$\sqrt{\frac{SSW_{O,T}(Y,Y)}{N_O + N_T}}$	$\frac{\hat{\sigma}_R^2}{2(N_O + N_T)}$
σ_E	$\sqrt{\frac{SSW_{P,TP}(Y - \hat{b}_X X, Y - \hat{b}_X X)}{N_P + N_{TP}}}$	$\frac{\hat{\sigma}_E^2}{2(N_P + N_{TP})}$

Table 5. Covariances for the new model.

$Cov(\hat{\theta}_1, \hat{\theta}_2)$ for $\theta_1 = a, b_T, b_P, b_{TP}$ and $\theta_2 = b_X, a, b_T, b_P$:

θ_2	b_X	a	b_T	b_P
θ_1				
a	0			(symm)
b_T	0	$-\frac{\hat{\sigma}_R^2}{N_0}$		
b_P	$\frac{\hat{\sigma}_E^2 \sum_{TP} X_i}{N_P SSW_{P,TP}(X_i, X_i)}$	$-\frac{\hat{\sigma}_R^2}{N_0}$	$\frac{\hat{\sigma}_R^2}{N_0}$	
b_{TP}	$\frac{(N_P + N_{TP})\hat{\sigma}_E^2 \sum_{TP} X_i}{N_P N_{TP} SSW_{P,TP}(X_i, X_i)}$	$\frac{\hat{\sigma}_R^2}{N_0}$	$-\frac{\hat{\sigma}_R^2}{N_0} - \frac{\hat{\sigma}_R^2}{N_T}$	$-\frac{\hat{\sigma}_E^2 \sum_{P,TP} X_i^2}{N_P SSW_{P,TP}(X_i, X_i)} - \frac{\hat{\sigma}_R^2}{N_0}$

$Cov(\sigma_R, \theta) = 0$, for $\theta = b_X, a, b_T, b_P, b_{TP}, \sigma_E$.

$Cov(\sigma_E, \theta) = 0$, for $\theta = b_X, a, b_T, b_P, b_{TP}, \sigma_R$.

Figure 1. A graphical depiction of the Solomon four-group design.

(O)	(R)			Y
(P)	(R)	X		Y
(T)	(R)		T	Y
(TP)	(R)	X	T	Y

Figure 2. The Solomon four-group design as a 2×2 factorial design.

	no pretest	pretest
no treatment	O	P
treatment	T	TP

Appendix 1: First derivatives

$$\frac{\partial \log L}{\partial b_X} = \frac{\sum_{P.TP} Y_i X_i - b_X \sum_{P.TP} X_i^2 - (a + b_P) \sum_{P} X_i - (a + b_P + b_T + b_{TP}) \sum_{TP} X_i}{\sigma_E^2}$$

$$\frac{\partial \log L}{\partial a} = \frac{1}{\sigma_R^2} \sum_{O.T} Y_i + \frac{1}{\sigma_E^2} \sum_{P.TP} Y_i - \frac{N_0 a}{\sigma_R^2} - \frac{N_T (a + b_T)}{\sigma_R^2} - \frac{N_P (a + b_P)}{\sigma_E^2} - \frac{N_{TP} (a + b_T + b_P + b_{TP})}{\sigma_E^2}$$

$$\frac{\partial \log L}{\partial b_T} = \frac{\sum_{T} Y_i}{\sigma_R^2} + \frac{\sum_{TP} Y_i}{\sigma_E^2} - \frac{b_X \sum_{TP} X_i}{\sigma_E^2} - \frac{N_T (a + b_T)}{\sigma_R^2} - \frac{N_{TP} (a + b_T + b_P + b_{TP})}{\sigma_E^2}$$

$$\frac{\partial \log L}{\partial b_P} = \frac{\sum_{P.TP} Y_i - N_P (a + b_P) - N_{TP} (a + b_T + b_P + b_{TP})}{\sigma_E^2}$$

$$\frac{\partial \log L}{\partial b_{TP}} = \frac{\sum_{TP} Y_i - b_X \sum_{TP} X_i - N_{TP} (a + b_T + b_P + b_{TP})}{\sigma_E^2}$$

$$\frac{\partial \log L}{\partial \sigma_R} = \frac{\sum_{O.T} Y_i^2 - 2a \sum_{O} Y_i - 2(a + b_T) \sum_{T} Y_i + N_O a^2 + N_T (a + b_T)^2}{\sigma_R^3} - \frac{(N_O + N_T)}{\sigma_R}$$

$$\frac{\partial \log L}{\partial \sigma_E} = \{ \sum_{P.TP} Y_i^2 - 2(a + b_P) \sum_{P} Y_i - 2(a + b_P + b_T + b_{TP}) \sum_{TP} Y_i + b_X^2 \sum_{P.TP} X_i^2 + 2b_X (a + b_P) \sum_{P} X_i + 2b_X (a + b_P + b_T + b_{TP}) \sum_{TP} X_i - 2b_X \sum_{P.TP} Y_i X_i + N_P (a + b_P)^2 + N_{TP} (a + b_T + b_P + b_{TP})^2 \} / \sigma_E^3 - (N_P + N_{TP}) / \sigma_E$$

Appendix 2: Second derivatives evaluated at the maximum

$$\frac{\partial^2 \log L}{(\partial b_X)^2} = -\frac{\sum_{P,TP} X_i^2}{\hat{\sigma}_E^2}$$

$$\frac{\partial^2 \log L}{\partial b_X \partial a} = -\frac{\sum_{P,TP} X_i}{\hat{\sigma}_E^2} = 0$$

$$\frac{\partial^2 \log L}{\partial b_X \partial b_T} = -\frac{\sum_{TP} X_i}{\hat{\sigma}_E^2}$$

$$\frac{\partial^2 \log L}{\partial b_X \partial b_P} = -\frac{\sum_{P,TP} X_i}{\hat{\sigma}_E^2} = 0$$

$$\frac{\partial^2 \log L}{\partial b_X \partial b_{TP}} = -\frac{\sum_{TP} X_i}{\hat{\sigma}_E^2}$$

$$\frac{\partial^2 \log L}{\partial b_X \partial \sigma_R} = 0$$

$$\frac{\partial^2 \log L}{\partial b_X \partial \sigma_E} = -2\frac{\sum_{P,TP} Y_i X_i - \hat{b}_X \sum_{P,TP} X_i^2 - (\hat{a} + \hat{b}_P)\sum_{P} X_i - (\hat{a} + \hat{b}_P + \hat{b}_T + \hat{b}_{TP})\sum_{TP} X_i}{\hat{\sigma}_E^3}$$

$$= -2\frac{\sum_{P,TP} Y_i X_i - \hat{b}_X \sum_{P} X_i^2 - \hat{b}_X \sum_{TP} X_i^2 - \frac{\sum_{P} Y_i \sum_{P} X_i}{N_P} + \frac{\hat{b}_X \sum_{P} X_i \sum_{P} X_i}{N_P}}{\hat{\sigma}_E^3}$$

$$-2\frac{\dfrac{\sum_{TP} Y_i \sum_{TP} X_i}{N_{TP}} + \dfrac{\hat{b}_X \sum_{Tp} X_i \sum_{TP} X_i}{N_{TP}}}{\hat{\sigma}_E^3}$$

$$= \frac{-2\left((SSW_{P.TP}(Y_i, X_i) - \hat{b}_X SSW_{P.TP}(X_i, X_i)\right)}{\hat{\sigma}_E^3} = 0$$

$$\frac{\partial^2 \log L}{(\partial a)^2} = -\frac{N_O}{\hat{\sigma}_R^2} - \frac{N_T}{\hat{\sigma}_R^2} - \frac{N_P}{\hat{\sigma}_E^2} - \frac{N_{TP}}{\hat{\sigma}_E^2}$$

$$\frac{\partial^2 \log L}{\partial a \partial b_T} = -\frac{N_T}{\hat{\sigma}_R^2} - \frac{N_{TP}}{\hat{\sigma}_E^2}$$

$$\frac{\partial^2 \log L}{\partial a \partial b_P} = -\frac{N_P}{\hat{\sigma}_E^2} - \frac{N_{TP}}{\hat{\sigma}_E^2}$$

$$\frac{\partial^2 \log L}{\partial a \partial b_{TP}} = -\frac{N_{TP}}{\hat{\sigma}_E^2}$$

$$\frac{\partial^2 \log L}{\partial a \partial \sigma_R} = \frac{-2\sum_{O.T} Y_i}{\hat{\sigma}_R^3} + \frac{2N_O \hat{a}}{\hat{\sigma}_R^3} + \frac{2N_T(\hat{a} + \hat{b}_T)}{\hat{\sigma}_R^3}$$

$$= \frac{-2\sum_{O.T} Y_i}{\hat{\sigma}_R^3} + \frac{2N_O \dfrac{\sum_O Y_i}{N_O}}{\hat{\sigma}_R^3} + \frac{2N_T \dfrac{\sum_T Y_i}{N_T}}{\hat{\sigma}_R^3} = 0$$

$$\frac{\partial^2 \log L}{\partial a \partial \sigma_E} = \frac{-2\sum_{P,TP} Y_i}{\hat{\sigma}_E^3} + \frac{2N_P(\hat{a}+\hat{b}_P)}{\hat{\sigma}_E^3} + \frac{2N_{TP}(\hat{a}+\hat{b}_T+\hat{b}_P+\hat{b}_{TP})}{\hat{\sigma}_E^3}$$

$$= \frac{-2\sum_{P,TP} Y_i}{\hat{\sigma}_E^3} + \frac{2N_P(\frac{\sum_P Y_i - \hat{b}_X \sum_P X_i}{N_P})}{\hat{\sigma}_E^3} + \frac{2N_{TP}(\frac{\sum_{TP} Y_i - \hat{b}_X \sum_{TP} X_i}{N_{TP}})}{\hat{\sigma}_E^3}$$

$$= \frac{-2\sum_{P,TP} Y_i}{\hat{\sigma}_E^3} + \frac{2\sum_P Y_i - \hat{b}_X \sum_P X_i}{\hat{\sigma}_E^3} + \frac{2\sum_{TP} Y_i - \hat{b}_X \sum_{TP} X_i}{\hat{\sigma}_E^3} = 0$$

$$\frac{\partial^2 \log L}{(\partial b_T)^2} = \frac{-N_T}{\hat{\sigma}_R^2} - \frac{N_{TP}}{\hat{\sigma}_E^2}$$

$$\frac{\partial^2 \log L}{\partial b_T \partial b_P} = \frac{-N_{TP}}{\hat{\sigma}_E^2}$$

$$\frac{\partial^2 \log L}{\partial b_T \partial b_{TP}} = \frac{-N_{TP}}{\hat{\sigma}_E^2}$$

$$\frac{\partial^2 \log L}{\partial b_T \partial \sigma_R} = \frac{-2\sum_T Y_i}{\hat{\sigma}_R^3} + \frac{2N_T(\hat{a}+\hat{b}_T)}{\hat{\sigma}_R^3} = \frac{-2\sum_T Y_i}{\hat{\sigma}_R^3} + \frac{2N_T \frac{\sum_T Y_i}{N_T}}{\hat{\sigma}_R^3} = 0$$

$$\frac{\partial^2 \log L}{\partial b_T \partial \sigma_E} = \frac{-2\sum_{TP} Y_i}{\hat{\sigma}_E^3} + \frac{2\hat{b}_X \sum_{TP} X_i}{\hat{\sigma}_E^3} + \frac{2N_{TP}(\hat{a}+\hat{b}_T+\hat{b}_P+\hat{b}_{TP})}{\hat{\sigma}_E^3}$$

$$= \frac{-2\sum_{TP} Y_i}{\hat{\sigma}_E^3} + \frac{2\hat{b}_X \sum_{TP} X_i}{\hat{\sigma}_E^3} + \frac{2N_{TP}(\frac{\sum_{TP} Y_i - \hat{b}_X \sum_{TP} X_i}{N_{TP}})}{\hat{\sigma}_E^3} = 0$$

$$\frac{\partial^2 \log L}{(\partial b_P)^2} = \frac{-N_P}{\hat{\sigma}_E^2} - \frac{N_{TP}}{\hat{\sigma}_E^2}$$

$$\frac{\partial^2 \log L}{\partial b_P \partial b_{TP}} = \frac{-N_{TP}}{\hat{\sigma}_E^2}$$

$$\frac{\partial^2 \log L}{\partial b_P \partial \sigma_R} = 0$$

$$\frac{\partial^2 \log L}{\partial b_P \partial \sigma_E} = -2\frac{\sum_{P.TP} Y_i - N_P(\hat{a}+\hat{b}_P) - N_{TP}(\hat{a}+\hat{b}_T+\hat{b}_P+\hat{b}_{TP})}{\hat{\sigma}_E^3}$$

$$= -2\frac{\sum_{P.TP} Y_i - N_P\left(\frac{\sum_P Y_i - \hat{b}_X \sum_P X_i}{N_P}\right) - N_{TP}\left(\frac{\sum_{TP} Y_i - \hat{b}_X \sum_{Tp} X_i}{N_{TP}}\right)}{\hat{\sigma}_E^3} = 0$$

$$\frac{\partial^2 \log L}{(\partial b_{TP})^2} = -\frac{N_{TP}}{\hat{\sigma}_E^2}$$

$$\frac{\partial^2 \log L}{\partial b_{TP} \partial \sigma_R} = 0$$

$$\frac{\partial^2 \log L}{\partial b_{TP} \partial \sigma_E} = -2\frac{\sum_{TP} Y_i - \hat{b}_X \sum_{TP} X_i - N_{TP}(\hat{a}+\hat{b}_T+\hat{b}_P+\hat{b}_{TP})}{\hat{\sigma}_E^3}$$

$$= -2\frac{\sum_{TP} Y_i - \hat{b}_X \sum_{TP} X_i - N_{TP}\left(\frac{\sum_{TP} Y_i - \hat{b}_X \sum_{TP} X_i}{N_{TP}}\right)}{\hat{\sigma}_E^3} = 0$$

$$\frac{\partial^2 \log L}{(\partial \sigma_R)^2} = -3\frac{\sum_{O.T} Y_i^2 - 2\hat{a}\sum_O Y_i - 2(\hat{a}+\hat{b}_T)\sum_T Y_i + N_O\hat{a}^2 + N_T(\hat{a}+\hat{b}_T)^2}{\hat{\sigma}_R^4}$$

$$+\frac{(N_O+N_T)}{\hat{\sigma}_R^2}$$

$$= -3\frac{\sum_{O,T} Y_i^2 - 2\frac{\sum_O Y_i \sum_O Y_i}{N_O} - 2\frac{\sum_T Y_i \sum_T Y_i}{N_T} + \frac{\sum_O Y_i \sum_O Y_i}{N_O} + \frac{\sum_T Y_i \sum_T Y_i}{N_T}}{\hat{\sigma}_R^4}$$

$$+\frac{(N_O + N_T)}{\hat{\sigma}_R^2}$$

$$= -3\frac{SSW_{O,T}(Y_i, Y_i)}{\hat{\sigma}_R^4} + \frac{(N_O + N_T)}{\hat{\sigma}_R^2} =$$

$$-3\frac{SSW_{O,T}(Y_i, Y_i)(N_O + N_T)^2}{\left(D_{O,T}(Y_i, Y_i)\right)^2} + \frac{(N_O + N_T)^2}{D_{O,T}(Y_i, Y_i)}$$

$$= \frac{-2(N_O + N_T)^2}{SSW_{O,T}(Y_i, Y_i)} = \frac{-2(N_O + N_T)}{\hat{\sigma}_R^2}$$

$$\frac{\partial^2 \log L}{\partial \sigma_R \partial \sigma_E} = 0$$

$$\frac{\partial^2 \log L}{(\partial \sigma_E)^2} = -3\frac{\sum_{P,TP} Y_i^2 - 2(\hat{a} + \hat{b}_P)\sum_P Y_i - 2(\hat{a} + \hat{b}_P + \hat{b}_T + \hat{b}_{TP})\sum_{TP} Y_i + \hat{b}_X^2 \sum_{P,TP} X_i^2}{\hat{\sigma}_E^4}$$

$$-3\frac{2b_X(\hat{a} + \hat{b}_P)\sum_P X_i + 2\hat{b}_X(\hat{a} + \hat{b}_P + \hat{b}_T + \hat{b}_{TP})\sum_{TP} X_i - 2\hat{b}_X \sum_{P,TP} Y_i X_i}{\hat{\sigma}_E^4}$$

$$-3\frac{N_P(\hat{a} + \hat{b}_P)^2 + N_{TP}(\hat{a} + \hat{b}_T + \hat{b}_P + \hat{b}_{TP})^2}{\hat{\sigma}_E^4} + \frac{(N_P + N_{TP})}{\hat{\sigma}_E^2}$$

$$= -3\frac{SSW_{P,TP}(Y_i, Y_i) + \hat{b}_X^2 SSW_{P,TP}(X_i, X_i) - 2\hat{b}_X SSW_{P,TP}(Y_i, X_i)}{\hat{\sigma}_e^4}$$

$$+\frac{(N_P + N_{TP})}{\hat{\sigma}_E^2}$$

$$=$$

$$-3\frac{\left(SSW_{P,TP}(Y_i, Y_i) + \hat{b}_X^2 SSW_{P,TP}(X_i, X_i) - 2\hat{b}_X SSW_{P,TP}(Y_i, X_i)\right)(N_P + N_{TP})^2}{\left(SSW_{P,TP}(Y_i, Y_i) + \hat{b}_X^2 SSW_{P,TP}(X_i, X_i) - 2\hat{b}_X SSW_{P,TP}(Y_i, X_i)\right)^2}$$

$$+\frac{(N_P+N_{TP})^2}{SSW_{P,TP}(Y_i,Y_i)+\hat{b}_X^2 SSW_{P,TP}(X_i,X_i)-2\hat{b}_X SSW_{P,TP}(Y_i,X_i)}$$

$$=\frac{-2(N_P+N_{TP})}{\hat{\sigma}_E^2}$$

Titles of Recent Research Reports from the Department of Educational Measurement and Data Analysis. University of Twente, Enschede, The Netherlands.

RR-99-06 G. van Engelenburg, *Statistical Analysis for the Solomon Four-Group Design*

RR-99-05 E.M.L.A. van Krimpen-Stoop & R.R. Meijer, *CUSUM-Based Person-Fit Statistics for Adaptive Testing*

RR-99-04 H.J. Vos, *A Minimax Procedure in the Context of Sequential Mastery Testing*

RR-99-03 B.P. Veldkamp & W.J. van der Linden, *Designing Item Pools for Computerized Adaptive Testing*

RR-99-02 W.J. van der Linden, *Adaptive Testing with Equated Number-Correct Scoring*

RR-99-01 R.R. Meijer & K. Sijtsma, *A Review of Methods for Evaluating the Fit of Item Score Patterns on a Test*

RR-98-16 J.P. Fox & C.A.W. Glas, *Multi-level IRT with Measurement Error in the Predictor Variables*

RR-98-15 C.A.W. Glas & H.J. Vos, *Adaptive Mastery Testing Using the Rasch Model and Bayesian Sequential Decision Theory*

RR-98-14 A.A. Béguin & C.A.W. Glas, *MCMC Estimation of Multidimensional IRT Models*

RR-98-13 E.M.L.A. van Krimpen-Stoop & R.R. Meijer, *Person Fit based on Statistical Process Control in an AdaptiveTesting Environment*

RR-98-12 W.J. van der Linden, *Optimal Assembly of Tests with Item Sets*

RR-98-11 W.J. van der Linden, B.P. Veldkamp & L.M. Reese, *An Integer Programming Approach to Item Pool Design*

RR-98-10 W.J. van der Linden, *A Discussion of Some Methodological Issues in International Assessments*

RR-98-09 B.P. Veldkamp, *Multiple Objective Test Assembly Problems*

RR-98-08 B.P. Veldkamp, *Multidimensional Test Assembly Based on Lagrangian Relaxation Techniques*

RR-98-07 W.J. van der Linden & C.A.W. Glas, *Capitalization on Item Calibration Error in Adaptive Testing*

RR-98-06 W.J. van der Linden, D.J. Scrams & D.L.Schnipke, *Using Response-Time Constraints in Item Selection to Control for Differential Speededness in Computerized Adaptive Testing*

RR-98-05 W.J. van der Linden, *Optimal Assembly of Educational and Psychological Tests, with a Bibliography*

RR-98-04 C.A.W. Glas, *Modification Indices for the 2-PL and the Nominal Response Model*

RR-98-03 C.A.W. Glas, *Quality Control of On-line Calibration in Computerized Assessment*

RR-98-02 R.R. Meijer & E.M.L.A. van Krimpen-Stoop, *Simulating the Null Distribution of Person-Fit Statistics for Conventional and Adaptive Tests*

RR-98-01 C.A.W. Glas, R.R. Meijer, E.M.L.A. van Krimpen-Stoop, *Statistical Tests for Person Misfit in Computerized Adaptive Testing*

RR-97-07 H.J. Vos, *A Minimax Sequential Procedure in the Context of Computerized Adaptive Mastery Testing*

RR-97-06 H.J. Vos, *Applications of Bayesian Decision Theory to Sequential Mastery Testing*

RR-97-05 W.J. van der Linden & Richard M. Luecht, *Observed-Score Equating as a Test Assembly Problem*

RR-97-04 W.J. van der Linden & J.J. Adema, *Simultaneous Assembly of Multiple Test Forms*

RR-97-03 W.J. van der Linden, *Multidimensional Adaptive Testing with a Minimum Error-Variance Criterion*

RR-97-02 W.J. van der Linden, *A Procedure for Empirical Initialization of Adaptive Testing Algorithms*

RR-97-01 W.J. van der Linden & Lynda M. Reese, *A Model for Optimal Constrained Adaptive Testing*

RR-96-04 C.A.W. Glas & A.A. Béguin, *Appropriateness of IRT Observed Score Equating*

RR-96-03 C.A.W. Glas, *Testing the Generalized Partial Credit Model*

RR-96-02 C.A.W. Glas, *Detection of Differential Item Functioning using Lagrange Multiplier Tests*

RR-96-01 W.J. van der Linden, *Bayesian Item Selection Criteria for Adaptive Testing*

RR-95-03 W.J. van der Linden, *Assembling Tests for the Measurement of Multiple Abilities*

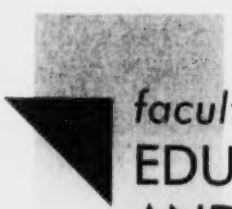
RR-95-02 W.J. van der Linden, *Stochastic Order in Dichotomous Item Response Models for Fixed Tests, Adaptive Tests, or Multiple Abilities*

RR-95-01 W.J. van der Linden, *Some decision theory for course placement*

RR-94-17 H.J. Vos, *A compensatory model for simultaneously setting cutting scores for selection-placement-mastery decisions*

...

Research Reports can be obtained at costs, Faculty of Educational Science and Technology, University of Twente, TO/OMD, P.O. Box 217, 7500 AE Enschede, The Netherlands.



faculty of
EDUCATIONAL SCIENCE AND TECHNOLOGY

A publication by
The Faculty of Educational Science and Technology of the University of Twente
P.O. Box 217
7500 AE Ens
The Netherla

BEST COPY AVAILABLE